

Gesture-Radar: a Dual Doppler Radar Based System for Robust Recognition and Quantitative Profiling of Human Gestures

Zhu Wang, *Member, IEEE*, Zhiwen Yu, *Senior Member, IEEE*, Xinye Lou, Bin Guo, *Senior Member, IEEE*, and Liming Chen, *Senior Member, IEEE*

Abstract—Gesture recognition is key to enabling natural human-computer interactions. Existing approaches based on wireless sensing focus on accurate identification of arm gesture types. It remains a challenge to recognize and profile the details of arm gestures for precise interaction control. In addition, current approaches have strict positioning requirements between radars and users, making them difficult for real-world deployment. In this paper, we adopt the multi-sensor approach and present Gesture-Radar - a dual Doppler radar-based gesture recognition and profiling system, which can capture subtle arm gestures with less positioning or environmental dependence. Gesture-Radar uses two vertically placed Doppler radars to collect complementary sensing data of gestures, based on which cross-analysis can be performed for gesture recognition and profiling. Specifically, we first propose a two-stage classification model and enhance the signal proximity matching method by applying constraint functions to the DTW algorithm, aiming to improve the accuracy of gesture type recognition. Afterwards, we establish and analyse unique features from the time-frequency spectrogram, which can be used to characterize in-depth gesture details, e.g., the angle or range of an arm movement. Experimental results show that Gesture-Radar achieves up to 93.5% average accuracy for gesture type recognition, and over 80% precision for profiling gesture details. This proves that the proposed approach is viable and can work in real-world environments.

Index Terms—Wireless sensing, gesture recognition, human-computer interaction, multiple sensors, Doppler radar, dual channel information.

I. INTRODUCTION

THE prevalence of the Internet of things and the rapid development of human recognition techniques have opened up opportunities for new ways of human and machine interaction. Traditionally if we want to control or interact with a certain equipment, like household appliances or industrial instrumentation, we need to use a console or a switch. This has been gradually and increasingly replaced by interactive controls. People prefer to use natural modalities of communication such as body movements to express their intentions or ideas, e.g., turning on or switching off a device by raising up or putting down their arms. In addition, it is necessary in some cases that users fine-tune their way of interaction in order to realize subtle controls, such as light intensity by

controlling the angle or range of the arm movement. To enable such natural human-computer interactions, a gesture sensing system needs to identify not only the types of arm movements but also microcosmic details, e.g., the angle or range of arm movements. Nevertheless, accurate recognition of a gesture still remains a challenge due to the complexity and diversity of human behaviors.

Currently, a number of technologies have been used to implement gesture recognition systems, e.g., wearable sensing [1]–[4], computer vision [5], [6], sound [7], [8], etc. While wearable device-based systems can achieve very high precision in detail, they are deemed as inconvenience in some occasions. Image processing-based recognition has privacy issues, and also requires good illumination conditions. For example, a user is not able to turn on a light by performing a certain gesture in a dark environment. As a mechanical wave, sound-based recognition approaches suffer from the drawbacks of severe attenuation of sound and limited effective range.

Due to the limitations of these aforementioned technologies, wireless sensing-based gesture recognition has recently attracted a growing attention. One of the most appealing approaches is based on Wi-Fi devices, which had been widely used for the recognition of gestures [9]. While it makes substantial progress, there is still several limitations, such as strict environment requirements and low stability. This leads to a number of open technical challenges, including (1) how to identify different gestures more accurately, (2) how to analyze the details of gestures, such as the angle or range of an arm movement, and (3) how to adapt such a system to different environments.

To address these challenges, we develop Gesture-Radar, an arm gesture recognition system based on multiple Doppler radar sensors, as shown in Fig. 1. The basic idea is to monitor a user with two Doppler radar sensors and then analyses the reflected waveforms. The radar sensors outputs dual channel signals called I and Q , containing information such as Doppler shift and signal intensity. Based on the principles of radar signal, Gesture-Radar can shield the interference from surrounding environments by extracting relevant information. Based on the collected signals, Gesture-Radar can use a phase detection method and a signal matching method to identify human gesture types from two radars respectively. It can then fuse the results to obtain a finalized cross-validated gesture type. Gesture-Radar also develops extra features to analyze the details of arm movements.

Z. Wang, Z. Yu, X. Lou and B. Guo are with the School of Computer Science, Northwestern Polytechnical University, Xi'an, 710072, China. E-mail: {wangzhu,zhiwenyu,guob}@nwpu.edu.cn

L. Chen is with the School of Computing, Ulster University, Newtownabbey, U.K. E-mail: l.chen@ulster.ac.uk

Manuscript received Dec. 18, 2019.

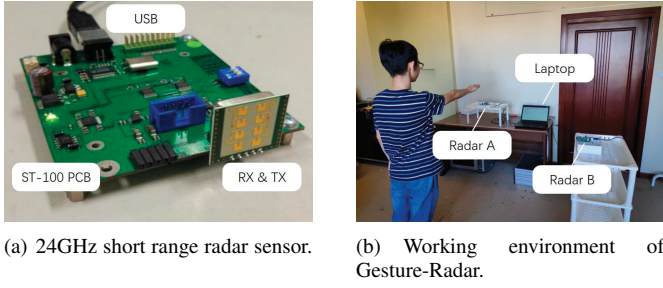


Fig. 1. Arm gesture recognition system.

To evaluate our approach, we select 12 different types of arm gestures for testing and experiment, as shown in Fig. 2, which are intuitively and widely used in activities of daily living. Our aim is not only to identify gesture types but also to capture the way these gestures are performed, i.e., the angle, range and completion level of arm movements. While accurate gesture recognition can be achieved with an extensive radar system (e.g., large radar arrays) usually along with high cost, Gesture-Radar has proved that it can achieve satisfactory recognition performance by using only two low-power radar sensors.

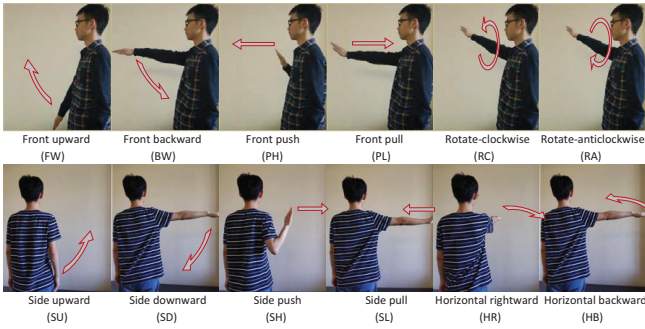


Fig. 2. Target gestures.

In this work, we have made the following contributions:

- **Robust gesture recognition using two radar sensors.** Gesture-Radar uses Doppler radar sensors, which are originally used for speed measurement, to capture and identify human gesture types and other details. In particular, to overcome shortcomings of the single radar-based system (e.g., signal blind area and environmental dependence), we leverage two vertically placed radars to collect complementary sensing data of gestures (i.e., both radial and tangential components) and propose corresponding methods to eliminate environment interference.
- **Accurate and quantitative profiling of gesture details based on multi-layer analysis.** In order to recognize gesture types accurately and efficiently, Gesture-Radar adopts a two-stage classification approach including a phase discrimination and a signal matching method. Meanwhile, to get more details about the gesture, we propose an efficient method to quantitatively estimate the angle or range of gestures by exploring and fusing the time-frequency spectrograms of two radar sensors.

- **Prototype implementation and performance evaluation.** To evaluate the implemented prototype, we recruited 10 volunteers in our experiment. Results illustrate that Gesture-Radar achieves an average accuracy of 93.5% for gesture type recognition, and can precisely profile the details of arm movements, e.g., the mean error of angle estimation is less than 15° (i.e., the precision exceeds 80%).

The remainder of this paper is organized as follows. Section II reviews related work. Section III provides some explanations about Doppler sensors and describes the system framework. We provide details about the recognition of gesture types and other related details in Section IV and Section V. Section VI presents experiments and results, followed by conclusion in the final section.

II. RELATED WORK

In general, we can divide existing works on wireless sensing-based behavior recognition into two categories: the contact device based approach and the non-contact device based approach. The first kind of approach requires the user to wear extra devices (e.g., RFID tags [1], [2], magnetic field sensors [3] or mobile devices [4]), which may bring inconvenience to the user during human-computer interactions. On the contrary, the second kind of approach (i.e., the contactless sensing approach) has the advantages of non-intrusiveness and scalability. In this section, we will focus the contactless sensing approach, which mainly consists of Wi-Fi based [9], [10] and radar based [11] approaches.

Wi-Fi based approach. During the past few years, Wi-Fi based human behavior recognition has become a popular research field, with a number of interesting applications. Early research uses Received Signal Strength (RSS) information to achieve localization, tracking or recognition [12]–[14]. However, as Wi-Fi signals are distributed unevenly in space, the recognition performance is relatively low.

Later, researchers shift their focus to Channel State Information (CSI), which contains detailed amplitude and phase information of different sub-carriers. By analyzing such information, CSI can be used for the recognition of fine-grained human behaviors, including mouth action [15], keystroke [16], gesture [17]–[19] and other behaviors [10], [20]–[26]. For example, WiFinger [18] was designed to identify digital gestures, and WiGeR [19] aimed to recognize palm movements. WiKey [16] can identify keystroke positions in a real world environments. In the Smokey system [21], authors proposed an indoor smoking monitoring method by mining the changing pattern of Wi-Fi CSI signals while someone smokes. By adopting different models, it is possible to recognize or monitor quite a number of human behaviors, such as indoor movements [26], respiration and heartbeats [10], [22], position [23], walking direction and falling [24], [25], and even human identity [9]. In addition to CSI, studies on the Doppler effect of Wi-Fi signals was also conducted, such as WiSee [27].

Although there have been great achievements in Wi-Fi based human behavior sensing and recognition, the inherent limitations of Wi-Fi devices cause difficulties to practical

applications. In particular, the Wi-Fi device's emission frequency is not fixed and its signal's spatial distribution is not uniform, which may lead to channel conflicts, large noise and packet dropouts [28]. Thereby, RSS, CSI and Doppler effect information of Wi-Fi signals are not stable, which raise strict requirements for device placement, space layout and user posture, making it difficult to deploy Wi-Fi based behavior recognition systems in real-world environments.

Radar based approach. In general, based on the hardware structure and the working principle, radar can be categorized into two types: the Frequency Modulation Continuous Wave (FMCW) radar and the Doppler radar. First, the transmitting frequency of FMCW radar usually varies in a linear manner over time, and it is easy to calculate the signal's round-trip time, based on which the distance between the radar and the target can be obtained [28]. Initially, lots of studies have been carried out to monitor and track targets using FMCW radars [29]–[31]. Afterwards, some subsequent efforts are made to identify human behaviors, e.g., gestures [32], [33] and activities [34]. However, while FMCW radars can be leveraged to recognize human behaviors, it is a challenging issue to obtain the angle or posture of human body's joints, resulting in poor performance.

Thereby, some recent studies choose to adopt a variety of sensors, such as the combination of FMCW radar and camera [32], while some other studies try to use Doppler radars for human behavior recognition [28], [35]–[39]. Traditionally, the Doppler radar is mainly used to detect moving objects. Different velocities and directions of moving objects can reflect different Doppler information, so it will perform well in recognition. For example, the RAM system [35] uses Doppler effect information to identify different human activities. Similarly, the Tongue-n-Cheek system [36] adopts Doppler radars to recognize the tongue's moving direction, allowing users to control the wheelchair with tongue movements. To sup up, most existing Doppler radar based studies focus on detecting the occurrence of a certain action or recognizing coarse-grained behaviors.

In contrast, this paper will investigate the possibility of recognizing more detailed information of behaviors by leveraging multiple Doppler radars. Our previous work [28], [39] used a single Doppler radar placed in front of the user to distinguish 5 different kinds of gestures. However, it can't obtain more detailed information, as the radar can only distinguish motion information in the radial component. Thus, this paper will address this issue by using two vertically placed radars, aiming to profile user behaviors in more details.

III. PRELIMINARIES AND SYSTEM OVERVIEW

Gesture-Radar is a wireless sensing system that utilizes the 24GHz K-LC2 short range radar sensors to enable natural human-computer interactions by recognizing arm gestures [28]. It leverages the property of Doppler shift [40], which denotes the difference between a wave's transmitted frequency and received frequency when the target moves relative to the observer / receiver. In particular, there will be a negative Doppler shift if the target moves away from the receiver, and a

positive Doppler shift if the target moves towards the receiver. Examples of Doppler shift are common in our daily life, e.g., different whistle will be heard when a train approaches or departs from the listener, and sonic boom effects can be produced by a supersonic aircraft.

The Doppler shift can be formally described as: if v is the velocity of a moving object, θ is the angle between the object's moving direction and the radial component of the radar, c is the light velocity which is much faster v (i.e., $c \gg v$), we can define Doppler shift as [28]:

$$f_d = f_r - f_t \approx \frac{2f_t \cdot v \cdot |\cos \theta|}{c}, \quad (1)$$

where f_t and f_r denote the frequency of the transmitted wave and the received wave, respectively.

In this work, we use a pair of dual channel Doppler radars, which output Doppler information with two orthogonal signals named I and Q that can be approximated as:

$$I(t) = \frac{A_R \cdot A_T}{2} \cdot \cos(2\pi \cdot f_d \cdot t + \varphi), \quad (2)$$

$$Q(t) = \frac{A_R \cdot A_T}{2} \cdot \sin(2\pi \cdot f_d \cdot t + \varphi), \quad (3)$$

where A_T and A_R represent the amplitude of the transmitted signal and the received signal, and φ denotes the original phase of the transmitted signal. We can find that there is a $\pi/2$ phase difference between I and Q . Accordingly, we can calculate the positive and negative values of f_d , and further analyze the direction of arm movements accordingly. Moreover, the dual channel signal contains amplitude, frequency and other information, which can be used to describe the target's relative moving velocity and angle. Therefore, Gesture-Radar is not only able to detect the occurrence of motions according to the mutation of Doppler signals, but also able to recognize micro-cosmic gesture details according to the changes of signals.

However, a problem of the Doppler radar is that it has a limited detection range and the Doppler shift will disappear if the target and the radar remain relatively static (e.g., θ equals $\pi/2$). In other words, the Doppler radar based sensing system will suffer from the problem of signal blind areas. For example, the radar placed in front of the user can't detect the side upward gesture. To solve this issue, we use two vertically placed radars to cover the environment and eliminate blind areas, and propose the corresponding gesture recognition model, as shown in Fig. 3. The framework can be characterized as a four-phase process, which is described in details as follows.

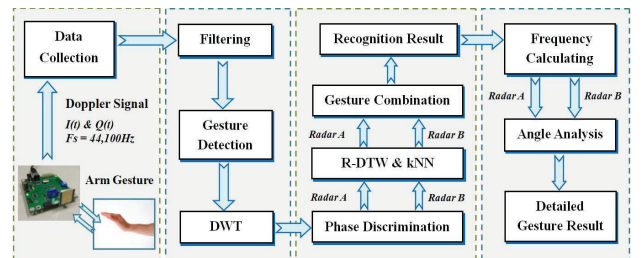


Fig. 3. Framework of the Gesture-Radar system.

In the first phase, dual channel signals are recorded with a pair of 24GHz radar sensors. Specifically, the radar located in front of the user is called “front radar”, and the radar on the right side of the user is called “side radar”. If someone performs a gesture, the dual channel information (I and Q) will be recorded by both radars and further uploaded to the server (a laptop in our experiment) with a sampling rate of 44,100Hz.

In the second phase, Gesture-Radar extracts motion segments by performing a series of signal processing procedures, and each segment corresponds to one user gesture. Gesture-Radar will first minimize the interference of environmental noise using a low-pass filter, and then uses a rule-based double-threshold comparison method to extract possible gesture segments. Finally, the Discrete Wavelet Transformation (DWT) method is adopted to compress the extracted gesture segments, so as to reduce the computation cost. As this part is similar to our previous work [39], we will not provide detailed descriptions in this paper.

In the third phase, Gesture-Radar makes use of the features of dual channel signals from two radars and adopts a two-stage classification method for gesture type recognition. Specifically, we first adopt a phase discrimination approach to calculate the direction of the gesture, and then identify the gesture with a signal matching method. In particular, the proposed signal matching method is composed of a Restricted Dynamic Time Warping (R-DTW) algorithm for waveform similarity calculation and a k-Nearest Neighbor (kNN) classifier for gesture recognition.

In the final phase, Gesture-Radar performs further analysis based on the result of gesture type classification. We mainly use a spectrum analysis method to obtain more details of arm movements. It includes two practical calculation methods: extremum ratio method and area ratio method. Gesture-Radar can finally get the angle, range or the cycle of gestures.

IV. GESTURE TYPE RECOGNITION

A. Target Gesture Set and Basic Idea of Gesture Type Recognition

After analyzing the main types of gestures that people use in daily life, we choose 12 typical gestures as the target gesture set for recognition, as shown in Fig. 2. Generally speaking, a gesture can be divided into three types from the perspective of any radar: *approaching*, *departing* and *uncertain*. For example, when a user performs *front upward* or *front push* gestures, the arm will move close to the front radar but almost has no effect on the side radar. When a user performs *horizontal rightward* or *horizontal backward* gestures, two radars will reflect the opposite direction of arm movements. When a *rotate* gesture is performed, the arm will neither move close to nor far away from the front radar, but we can get the direction information from the side radar. The above observations play an important role in the development of the Gesture-Radar system. In Table I, we list gestures of different motion directions to each radar.

In addition, some gestures are considered to be similar for a particular radar. For example, when a user performs the *front*

TABLE I
GESTURE DIRECTIONS TO EACH RADAR.

Case	Front Radar	Side Radar
approach	front upward	side upward
	front push	side push
	horizontal backward	horizontal rightward
depart	front backward	side downward
	front pull	side pull
	horizontal rightward	horizontal backward
uncertain	other gestures	other gestures

upward or *horizontal backward* gestures, the front radar cannot distinguish them theoretically. Because these two gestures are identical to the radial component of the front radar, and they can only be further classified by the side radar. According to this observation, the 12 target gestures can simply be divided into 6 cases, including *upward* (means the signal is similar to a front upward gesture observed by the front radar, the same below), *downward*, *push*, *pull*, *rotate* and *unrecognizable* (means the radar can not classify a specific case, e.g., the side radar is not sensitive to a *front upward* gesture as f_d is very low). Each radar only needs to classify the gesture into one of the cases, and Gesture-Radar will get the final gesture types by fusing the recognition results from the two radars. In such a way, the number of labels for classification are reduced, and the gesture recognition accuracy and efficiency can be greatly improved.

In summary, Gesture-Radar adopts a two-stage classification approach to recognize different types of gestures, as shown in Fig. 4. First, at the gesture direction identification stage (Stage I), the system identifies whether a gesture is moving close or far away from any radars based on phase discrimination. At the gesture type classification stage (Stage II), gestures are further classified into 6 cases from the perspective of any radars based on R-DTW and kNN. Afterwards, we fuse the results from two radars to obtain the final recognition result. Specifically, if Gesture-Radar fails to identify a gesture’s moving direction at Stage I, the system will compare it with all the gesture waveforms at Stage II.

B. Gesture Direction Identification Based on Phase Discrimination

Based on Doppler principles, we can obtain the direction of a gesture through Doppler shift. The basic idea is that the dual channel signals of the two radar sensors are orthogonal, and thus we can obtain their phase difference based on the value of f_d . In Fig. 5, we show the phase information of a *front forward* gesture and a *front backward* gesture observed by the front radar. In case that the target is moving towards the radar, i.e., $f_d > 0$, the I signal will have a phase advance of about $\pi/2$. On the contrary, the Q signal will have a $\pi/2$ phase advance if the target is moving away from the radar. Therefore, based on Equations 2 & 3, a signal’s phase can be calculated as follows:

$$\Theta = \arctan \frac{Q(t)}{I(t)} = 2\pi \cdot f_d \cdot t + \varphi = \frac{4\pi \cdot v \cdot \Delta t}{\lambda}. \quad (4)$$

The changing rate of Θ reflects a gesture’s relative moving direction to the radar. However, it is difficult to compute Θ

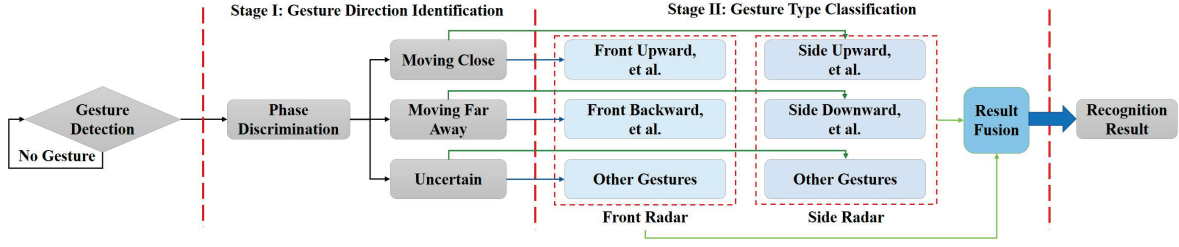


Fig. 4. Framework of the two-stage gesture classification method.

directly, as the function \arctan is only continuous within $(-\pi/2, \pi/2)$. Therefore, to correctly estimate the phase, Gesture-Radar needs to eliminate discontinuous points and reconnect them by performing phase unwrapping. As shown in Fig. 5, the unwrapped phase of an approaching gesture has an upward trend and that of a departing gesture has a downward trend, which is the embodiment of Doppler information. Accordingly, we can determine the direction of gestures by applying a head-tail subtraction method with a threshold parameter Δp . In particular, if the head-tail difference of \arctan is higher than $|\Delta p|$ or lower than $-|\Delta p|$, the gesture will be identified as approaching or departing, respectively. Otherwise, i.e., the difference is within $(-|\Delta p|, |\Delta p|)$, the gesture will be recognized as uncertain, indicating that the system is not capable of determining gesture direction at Stage I. Based on the theoretical analysis, the phase discrimination has a high accuracy in analyzing the direction of arm movements, which help us avoid possible error choices at Stage II. To sum up, Stage I can improve the overall classification performance, as it gets rid of some computation cost.

C. Gesture Similarity Assessment with Restricted Dynamic Time Wrapping

The received signal of a Doppler radar is influenced by the distance and angle between the user and the radar, as well as the way a gesture is performed, e.g., the strength, the speed, and the range. For example, a faster action will produce a larger Doppler shift, thus resulting in a higher signal frequency. If the speed of a gesture is not constant, the frequency of the received signal will change as well, which makes it difficult to quantify the Doppler shift. Meanwhile, the distance between a user and a radar will also affect the amplitude of the received signal. In other words, it is difficult to characterize gestures by extracting features from the received signal of Doppler radars, as the values of such features are quite unstable. In our experiments, we have found that feature-based classifiers, e.g., Naive Bayes, Random Forest, and Support Vector Machine, not only involve complex feature extraction process but also suffer from high classification errors in the context of different users and environments. Nevertheless, we also discovered that while the features of various gestures of the same type might be quite different, the overall changing trends of the corresponding waveforms are very similar. According to such an observation, Gesture-Radar adopts a signal matching method and a kNN classifier to classify different arm gestures.

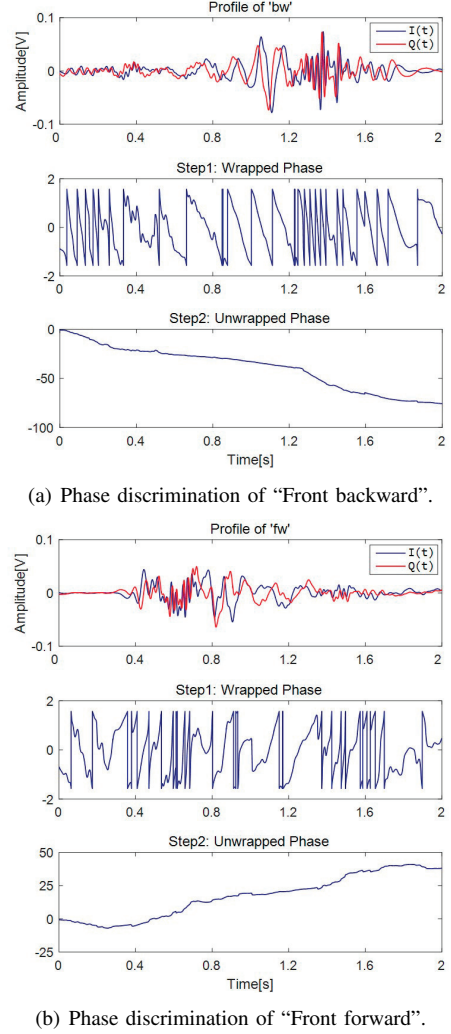


Fig. 5. The process of phase discrimination.

In general, the performance of a kNN classifier mainly depends on the adopted similarity measure. Specifically, if a pair of gesture waveforms have different lengths or speeds, most measures (e.g., Euclidian) will not be able to quantify their similarity accurately, leading to high errors. Therefore, we adopt the Restricted Dynamic Time Wrapping (R-DTW) algorithm proposed in our previous work [28], which restricts the matching paths of the original Dynamic Time Wrapping algorithm [41] to calculate the distance between two gesture waveforms.

In particular, given a pair of gesture waveforms $X =$

(x_1, x_2, \dots, x_m) and $Y = (y_1, y_2, \dots, y_n)$ (m and n denote the lengths of waveforms), DTW first introduces a matrix $d(m, n)$ to record the distance between any pair of points in X and Y as follows:

$$pd(i, j) = [x(i) - y(j)]^2. \quad (5)$$

Afterwards, a new matrix $D(m, n)$ is constructed to represent the overall distance, where each entry $D(i, j)$ is defined as follows:

$$D(i, j) = \min \left\{ \begin{array}{l} [d(i, j) + D(i-1, j)], \\ [d(i, j) + D(i, j-1)], \\ [2d(i, j) + D(i-1, j-1)] \end{array} \right\}. \quad (6)$$

Based on dynamic programming, the final distance $D(m, n)$ between two waveforms can be obtained, and the optimal warping path can be restored. However, the original DTW algorithm has some shortcomings. Ideally, the optimal matching path of two same type gestures should be close to the diagonal line (a straight line from $D(1, 1)$ to $D(m, n)$) [28], and the calculation of some corner points is meaningless. In other words, it is not appropriate to conclude that two waveforms belong to the same type of gestures if their optimal matching path needs to go through points that are far away from the diagonal line. Thereby, we introduce boundary functions to restrict the calculation process of DTW, and define the revised method as R-DTW. Consequently, the function will divide the whole matrix into three parts, and only points from the middle part can appear in the matching path. Because different functions may lead to different results, we introduce three different constraint boundary functions, as shown in Fig. 6), which are linear constraint, rhombic constraint and parabolic constraint.

R-DTW can not only process waveforms of different lengths but also address the problem caused by most similarity measures (e.g., the Euclidean distance). Moreover, as the possible matching path is restricted, R-DTW also improves the matching accuracy and reduces the computation overhead.

D. Gesture Type Classification and Interference Minimizing

According to the core idea of the kNN algorithm, we know that its performance mainly depends on the most adjacent samples rather than the whole data set. Thereby, compared with other classification methods, kNN is more suitable for overlapping samples or domains. In this work, we adopt R-DTW as the distance measure to build the kNN classifier, and set the value of k as 7 based on empirical results. We first let each radar obtain a classification conclusion based on the two-stage classification method [28], and then fuse the conclusions of the two sensors based on rules and get the final recognition result, as shown in Table II. Though the rule is intuitive, to help better understand the design rationale, we provide additional explanations as follows.

- Case 1: The trend of arm motion direction detected by the side radar is: approach→depart→approach.
- Case 2: The trend of arm motion direction detected by the side radar is: depart→approach→depart.

TABLE II
GESTURE TYPE CLASSIFICATION RULES.

Front Radar	Side Radar	Recognition Result
upward	unrecognizable	FW
downward	unrecognizable	BW
push	unrecognizable	PH
pull	unrecognizable	PL
rotate	Case 1	RC
rotate	Case 2	RA
upward	upward	FW
downward	downward	BW
unrecognizable	upward	SU
unrecognizable	downward	SD
unrecognizable	push	SH
unrecognizable	pull	SL
downward	upward	HR
upward	downward	HB
push	push	PH
pull	pull	PL

First, there is no “unrecognizable” gesture samples in the training set. Although *unrecognizable* is a possible classification result, it does not simply mean that the radar fails to recognize a gesture, but the radar finds no relationship between the gesture and itself. Based on the results of extensive experiments, we add additional rules to identify *unrecognizable* gestures. On one hand, during the gesture detection phase, as a radar can only capture the motion information on the radial components, there might be cases that one radar captures a gesture while the other one fails, e.g., the user performs a gesture such as *front upward* or *side push*. In such cases, we will directly define the classification result of the radar which didn’t capture a gesture as *unrecognizable*. On the other hand, during the gesture type classification phase, there might be cases that the average R-DTW distance between the target gesture and samples of any gesture types in the training set is larger than a predefined threshold, which means that the target gesture is not similar to any gestures. In such cases, the classification result will also be set as *unrecognizable*.

Second, if two radars obtain the same classification result (e.g., both are *upward*), it means that this is a “non-standard gesture” which needs to be profiled in more depth to get the detailed information. We will explain it in Section V.

Third, we need to take into account some complex factors when designing the Gesture-Radar system, such as the multi-person environment. Thereby, Gesture-Radar should be able to distinguish interference behaviors (e.g., drinking water, using mobile phones, and shaking body) caused by other users and shield the response. Since an interference behavior usually has no regularity, we define the following rule: if the combination result of two radars does not exist in Table II, Gesture-Radar will conclude that it is an interference behavior and ignore it directly.

V. GESTURE DETAIL PROFILING

In order to achieve natural human-machine interactions, only knowing gesture types is not enough. For example, users will not always raise their hands exactly to the front (*front upward*) or the right side (*side upward*) of their body. Naturally, they may raise hands in the direction of any angle between the front and the right side, and so do the *push* and

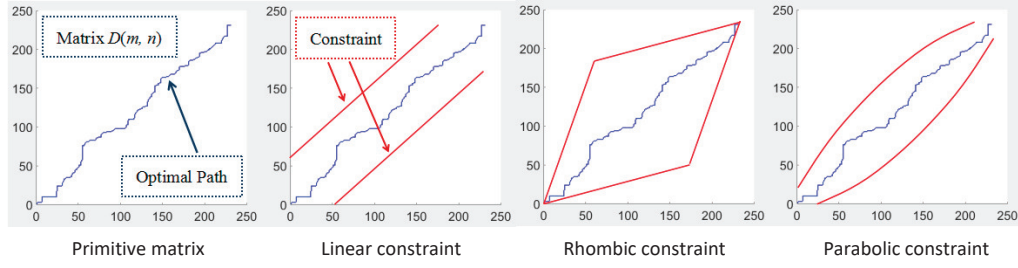


Fig. 6. Constraint functions of R-DTW. **Linear Constraint:** $(n/m)x - an \leq y \leq (n/m)x + an$, $0 < a < 1$; **Rhombic Constraint:** The slopes are k and $1/k$ ($k > 1$), and the curves pass through points $(km/(k+1), n/(k+1))$ and $(m/(k+1), kn/(k+1))$; **Parabolic Constraint:** $(kn/m^2)x^2 + (1-k)nx/m - a \leq y \leq (-kn/m^2)x^2 + (1+k)nx/m + a$, $0 < k < 2$, $a > 0$.

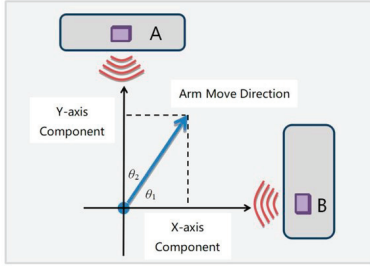


Fig. 7. Quantitative gesture profiling with a coordinate system.

the *pull* gestures. In addition, there are also cases that the user performs incomplete *horizontal rightward* or *horizontal backward* gestures, because it is common for an arm to swing in a small area in front of the body. We call such gestures as “*non-standard gestures*”, which only change the information in the radial component of each radar and have little effect on the overall shapes of the gesture waveforms. Therefore, a signal matching method can only distinguish what gestures had been performed, but can’t obtain the details (e.g., the angle or range of arm movements). This section will address the above issue by developing a quantitative gesture detail profiling method.

A. Signal Transformation Based on Spectrum Conversion

We build a coordinate system from the top view of a scene, as shown in Fig. 7, where the X axis points to the radar on the right side of the body, and the Y axis points to the radar in front of the body. These two axes are exactly perpendicular.

If a user raises her arm in the direction of angle θ , the ratio of the speeds detected by the two radars at any time should satisfy the following formula:

$$\frac{v_A}{v_B} = \tan \theta_1 = \cot \theta_2. \quad (7)$$

Details of any other gestures can be obtained in the same way. For example, if we want to get the range of a horizontal movement, we only need to calculate the speed ratio at the end of a gesture. Thus, it is easy to restore the value of θ as long as we can calculate the speeds of an arm movement by exploring the signal received by the two radars.

The Doppler radar is traditionally used to measure the speed of a moving object as a whole. However, when a user performs a certain arm gesture, the actual speeds of different parts of the arm might be different, thus it is very difficult

to calculate the speed. Therefore, Gesture-Radar uses the frequency of I and Q signals, which represent the Doppler shift, to indirectly analyze the velocity of arm movements. Specifically, we adopt a spectral analysis method to transform the time-amplitude waveforms of I and Q signals into time-frequency spectrograms, as shown in Fig. 8. Based on such a transformation, we can obtain the relationship between the frequency and the time domains of a signal. Areas where the color is more red represent that the corresponding frequency has a higher signal intensity at that time point. We analyze the time-frequency spectrogram of three similar gestures (users raise their arms to slightly different directions), and the results show that there are significant differences among the received signals in the frequency domain, which is an effect that previous signal matching method cannot achieve.

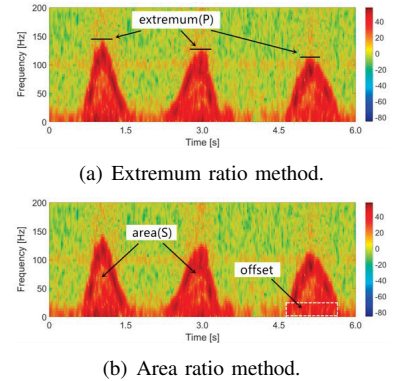


Fig. 8. Time-frequency spectrograms of 3 slight different gestures captured by the front radar (The angle θ_2 from left to right are 0 , $\pi/12$, and $\pi/6$).

B. Gesture Detail Profiling Based on Time-frequency Spectrogram

As the speeds of different arm parts differ from each other, a large amount of low frequency information will appear at the same time, i.e., red areas in the spectrogram, which represent useful gesture information. For example, as shown in Table III, we present some statistics (e.g., mean/maximum/minimum peak-frequency/area) of the spectrogram for different gestures, which are obtained based on 400 gestures from the same participant (i.e., 100 gestures for each of the 4 types). We can find that there are significant differences among the statistical measures of different gesture types.

TABLE III
STATISTICS OF SPECTROGRAM FOR DIFFERENT GESTURES.

Gesture Type	Peak-Frequency			Area		
	Mean	Max	Min	Mean	Max	Min
FW	311.3	506.0	215.3	8018.1	10188.8	6192.8
PH	247.4	398.4	183.0	5917.8	10704.3	4770.1
PL	153.3	193.8	118.4	4954.2	11592.5	1003.0
RA	163.2	258.4	107.7	10409.3	16321.4	7466.9

Thereby, to obtain the details of gestures, we propose two methods for gesture angle and range calculation. The first method, as shown in Fig. 8(a), is to directly compare the extreme value of the spectrum of each gesture using the following formula:

$$\theta_1 = \arctan\left(\frac{P_A}{P_B}\right), \quad (8)$$

$$\theta_2 = \arctan\left(\frac{P_B}{P_A}\right), \quad (9)$$

where P_A and P_B represent the extreme values of the two spectrograms. The basic idea is that the ratio of the arm movement speeds observed by two radars should be the same at any moment, thus we only use the extreme values for calculation. Though this method is intuitive, there is a problem when calculating the edge angle. For example, even if the arm is raised to the exact front, the spectrogram of the side radar will still produce a small amount of low frequency signals (usually around 20Hz), and there will be a deviation of about 15~20 degrees when using the above formula.

The second method is to use the respective area of the spectrum produced by two radars. While it is more complicated than the first method, it is able to characterize the whole gesture more accurately, thus leading to an improved performance. Specifically, the definition is as follows:

$$\theta_1 = \arctan\left(\frac{S_A - offset}{S_B - offset}\right), \quad (10)$$

$$\theta_2 = \arctan\left(\frac{S_B - offset}{S_A - offset}\right), \quad (11)$$

where S_A and S_B represent the corresponding areas of the two spectrograms, respectively. Considering the precision deviation caused by the ultra-low frequency part (below 20Hz), we can correct it by introducing an offset which is defined as the area within the rectangle, as shown in Fig. 8(b). For a spectrogram, as the offset only occupies a small part of the area, the overall calculation will not be significantly affected. Nevertheless, the result will be greatly improved when calculating an edge angle.

Both the above two methods have their advantages, and we will show the experimental results in Section VI-C. Obviously, based on time-frequency analysis, Gesture-Radar can quantitatively profile arm movements in more details.

VI. EXPERIMENTAL RESULTS

A. Experiment Setup and Data Collection

To evaluate the performance of Gesture-Radar, we use two 24GHz K-LC2 short range radars with the ST-100 radar development board and a laptop with dual core CPU and 8GB

ROM. As shown in Fig. 1(b), both radars are placed at a height of 1.4m above the ground while a user stands about 1.5m away. During experiments, radars continuously send signals at the rate of 24GHz. Meanwhile, the development board transforms the captured echo signal into dual channel data, and uploads the data to the laptop at a sampling rate of 44,100Hz [28].

The data collection process consisted of two parts. In the first part, we collected data from 10 volunteers, who first performed “*standard gestures*” (i.e., in the specified order) and then *free-style gestures* (i.e., in any order they want, e.g., *front forward*, *horizontal rightward*, *side pull*, and *side push*). Participates repeated each gesture for 10 times with an average time interval of about 1~2s. In general, each volunteer contributed 120 gesture samples and each sample contains I and Q dual channel time domain signals. Specifically, we adopted 10-fold cross validation to verify the accuracy of gesture type recognition.

In the second part, volunteers were asked to perform “*non-standard gestures*” with any angle or range of arm movements, and each user also provided 120 gesture samples. Afterwards, we used the data collected in the first part as a training set to build the gesture type recognition model, based on which we classified the gesture samples collected in the second part. Then, we further used the frequency analysis method to obtain the details such as angle or range of these arm movements.

B. Gesture Types Recognition Performance

1) *Gesture direction identification accuracy*: Based on the findings in our previous work [28], [39], threshold $\Delta p = 0$ is enough and the detection accuracy cannot be improved by increasing the threshold. Therefore, the parameter Δp is set as zero during the experiment. The performance of direction identification is summarized in Table IV. Specifically, as Δp is 0, there is no classification result of *uncertain* in Table IV, where *uncertain* means Gesture-Radar failed to detect the gesture, and the recognition result would be *unrecognizable*.

TABLE IV
DIRECTION DETECTION ACCURACY OF DIFFERENT TYPES OF GESTURES.

	Front Radar		Side Radar	
	Direction	Accuracy	Direction	Accuracy
FW	approach	99%	uncertain	96%
BW	depart	100%	uncertain	98%
SU	uncertain	99%	approach	100%
SD	uncertain	98%	depart	99%
HR	depart	100%	approach	100%
HB	approach	100%	depart	100%
PH	approach	97%	uncertain	94%
PL	depart	100%	uncertain	93%
SH	uncertain	89%	approach	100%
SL	uncertain	95%	depart	100%
RC	56% as approach, 44% as depart		/	
RA	59% as approach, 41% as depart		/	

We can see that the average accuracy of the first 10 gesture types is about 97.9%, while that of the other 2 gesture types is outlier, no matter which direction the user rotates. It means that the front radar can not capture the information of rotation based on phase discrimination. In other words, in order to identify a *rotate gesture*, each gesture in Gesture-Radar need be compared with rotate waveforms based on the

R-DTW distance. Specifically, if a gesture was identified as *rotate* based on the follow-up signal matching method, we can analyze whether it is clockwise or anti-clockwise by the phase change trend using the side radar. As shown in Table I, the phase discrimination procedure is able to eliminate many possible incorrect choices.

2) *Gesture type classification accuracy*: To validate the system's performance on gesture classification, we tried different k as well as different constraint boundary functions of R-DTW. Experimental results indicate that the optimal performance was obtained when the number of nearest neighbor is set as 7 and the parabolic constraint function (with parameters $k = 0.67$ and $a = 10$) was used, as shown in Fig. 9.

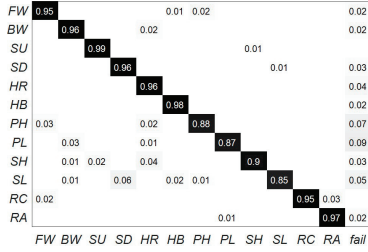


Fig. 9. Recognition accuracy of different types of gestures.

In total, Gesture-Radar achieved an average recognition accuracy of about 93.5% and a F1-score of 95.0%. Meanwhile, we also present the ROC curve of different gesture types 10. According to these measures, we can conclude that the proposed R-DTW & kNN approach achieves high performance. However, we found that a small number of gestures had not been classified correctly, while another small number of gestures failed to be classified as any types. This is because some gestures (i.e., *front forward*) only affect one of the two radar sensors, and if the affected radar failed to detect the gesture, the whole system will not be able to determine which gesture the user had actually performed. Generally speaking, Gesture-Radar has a strong adaptability during the experiment as all the volunteers obtained good human-computer interaction experience.

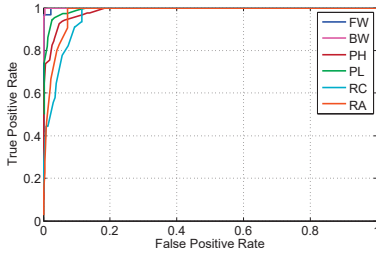


Fig. 10. ROC curve of different types of gestures.

Moreover, we compared our experimental results with the RAM system [35], which also used two Doppler radars to recognize human activities. Specifically, we extracted 6 features in the same way as the RAM system, including average, standard deviation, range, total energy, phase difference and velocity. Then, we tested different classification models, including Naive Bayes (NB), Random Forest (RF) and

Support Vector Machine (SVM) by performing 10-fold cross validation, as shown in Fig. 11. In particular, we replaced the R-DTW & kNN module in the proposed framework with the combination of a feature extraction module and a classification module, i.e., other modules of these baseline solutions are the same as the proposed framework.

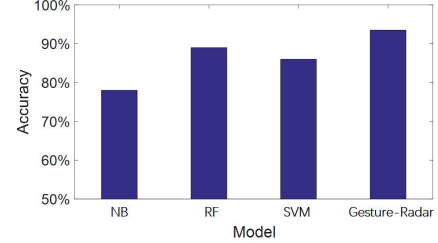


Fig. 11. Average accuracy of different recognition models.

It is observed that the average recognition accuracy of the two-stage classification model proposed in Gesture-Radar is higher than that of the classical classification models in the same experimental environment. The RAM system aims to recognize human activities such as dancing, walking, washing and other coarse-grained behaviors. Therefore, the feature extraction based model can work well. However, in order to distinguish subtle behaviors such as arm movements, the accuracy will decrease severely. In summary, our proposed system has a higher arm gesture recognition ability.

Further, to validate the proposed rule-based sensor fusion mechanism, we compared it with another two baseline mechanisms, as shown in Table V. The first one is *Max.*, which means that the recognition result of the dual-radar system depends on the sensor with higher matching probability. The second one is *Avg.*, which denotes that the result is obtained based on the average matching probability of the front and side radars, i.e., the recognized gesture type is the one with the highest average matching probability.

TABLE V
PERFORMANCE OF DIFFERENT SENSOR FUSION MECHANISMS.

	Max.	Avg.	Rule-based
FW	95%	84%	95%
BW	95%	85%	96%
SU	94%	82%	99%
SD	92%	80%	96%
HR	88%	76%	96%
HB	90%	78%	98%
PH	85%	74%	88%
PL	84%	75%	87%
SH	88%	76%	90%
SL	82%	74%	85%
RC	75%	70%	95%
RA	78%	72%	97%

Accordingly, we can find that the proposed rule-based sensor fusion mechanism outperforms the two baseline mechanisms, especially for complicated gestures such as *rotates*. The reason is that it has leveraged the logical relationship of the two radars, which are vertically placed in the environment. We also find that the performance of the *Avg.* mechanism is worse than that of the *Max.* mechanism, which should be due

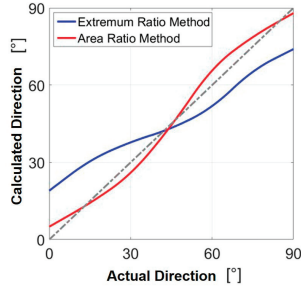


Fig. 12. Gesture details profiling results.

to the fact that in most cases only one sensor can capture valuable information when a certain gesture is performed, and the sensor with poor information usually outputs quite low matching probabilities, making the average probability of two sensors less accurate. To sum up, the rule-based sensor fusion mechanism is more suitable for the dual radar system.

C. Gesture Details Profiling Performance

Since users can raise their arms in any angle or move arms horizontally to any direction, in the second part of the data collection, we used cameras to record the actual moving angle or range of the user's "non-standard gestures". Specifically, while a "non-standard gesture" affects the Doppler shift and the signal intensity in a different manner, it still has similar overall waveform shape as the standard gesture. Thereby, the recognition accuracy of gesture types is still as high as 92.4%. Gesture-Radar further analyzes details of these successfully recognized gestures, and the results are shown in Fig. 12.

We define the recognition precision as $1 - (\text{average error of angle} / 90^\circ)$. Obviously, the profiling results are strongly correlated with the actual direction of arm movements, and the area ratio method has a higher average precision (more than 80%), which means that we can profile gesture details in a quantitative manner. It is noteworthy that the average errors of the two methods are both relatively low when the actual angle or range is around 45° . This is due to the fact that the Doppler shift components observed by the two radars are quite similar in such cases. In addition, we can found that the shortcoming of the extremum ratio method at the edge region (i.e., near 0° and 90°) can be well solved by the area ratio method. In the future, we plan to further improve the recognition accuracy by improving the angle estimation algorithm.

D. Robustness of the System

1) *Room Layout and Device Positioning*: We performed experiments in different environment layouts, including 3 rooms (a bedroom, a living room and a conference room) and 6 deployment positions by varying the device's height and distance, as shown in Table VI.

TABLE VI
CASES OF DIFFERENT ENVIRONMENT LAYOUTS.

Case	Device Height	User Distance
1	1.2m	1.5m
2	1.4m	1.5m
3	1.2m	3m
4	1.4m	3m
5	1.2m	4m
6	1.4m	4m

Experimental results are shown in Fig. 13. On one hand, we can see that the system's performance was mainly influenced by the distance between the target and the radar. Specifically, Gesture-Radar performs well while the distance is less than 3 meters, and the accuracy would decrease significantly once the distance increases any longer. The reason is that the echo signal from a distant target becomes weak, making it difficult to detect and recognize gestures. On the other hand, there is no significant difference between the average performance of 3 different rooms, and the performance change is very slight when the height of the radar varied from 1.2 to 1.4 meters. Thus, we can conclude that the distance between the user and the radar is particularly important for system.

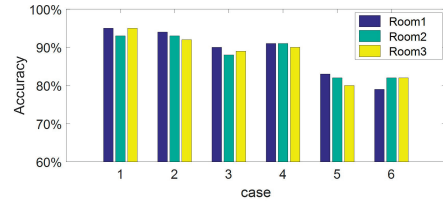


Fig. 13. Gesture type recognition accuracy in different environment.

2) *Sampling Rate of the Doppler Radar*: Sampling rate is an important factor that influences the system's performance, as the echo signal will contain more detailed information if a higher sampling rate is applied. The radar we used is deployed on a commercial ST-100 development board, which uploads data at a fixed sampling rate of 44,100Hz. To test the influence of different sampling rates, we chose to vary the sampling rate of the computer's sound card. Results shown that, if the sampling rate is set to 22,050Hz, the average recognition accuracy of one user's gestures reduced from 95.4% to 93.3%. In case that the sampling rate is further cut down to 4,410Hz, the gesture type recognition accuracy is only about 88% while the gesture detail profiling precision decreases to 60%. Obviously, Gesture-Radar needs a high sampling rate to achieve satisfactory performance.

3) *Multi-Person Environment*: We tested the system's recognition performance in multi-person (2~4 users) environments, where only one of them is the target user. Specifically, when the target subject performs gestures, the non-target ones were asked to perform some interfering actions at different locations. We observed that if there was someone who performs slight interference actions (e.g., shaking the head) at 3 meters away from the radars or keeps quiet (e.g., reading a book) within 3 meters, the system's performance will not significantly influenced. For example, in a 4-person room, the recognition accuracy of Gesture-Radar is still above 90% and

the precision of gesture detail profiling is higher than 75%. However, if a non-target user performs big range movements (e.g., jumping or running), Gesture-Radar will fail to work.

E. Discussion

1) *Synchronization of Multiple Radar Sensors*: When it comes to the use of multi-radar sensors, we have to face the problem of synchronization. Traditional radar arrays require a very high synchronization rate of less than $1\mu\text{s}$ [34]. For our system, it will result in a very high cost to control the radar transceivers. In fact, we have considered this problem when designing the system. Gesture-Radar chose to let the two radars obtain their own classification results and then fuse the results at the information center (the laptop in our experiment), so the two radars can work independently. A remaining issue is the detection of gestures, thereby we designed a buffer for each radar to store the latest signal. In particular, we set the buffer size as 105.8K (i.e., $44,100\text{Hz} \times 2.4\text{s}$), since a gesture will not exceed 1.2s in most cases. The gesture detection algorithm will run based on the data in the buffer. Once a completed signal is detected, Gesture-Radar will recognize it as a gesture. At the same time, it will immediately send information to the other radar buffer to get the detection result. Experiments show that the time cost of this step is about 0.1s, which means there might be a time offset of 0.1s in the signal segments detected by the two radars. Although the radars do not work synchronously, the system will not be affected, as the subsequent classification process is independent.

2) *Environment Dependence*: Environment dependence is one of the major challenges of most existing RF-based sensing systems. To mitigate this problem, we leverage two radar sensors to collect complementary information for more robust behavior recognition. However, the environment (e.g., location and orientation) dependence problem remains. Fortunately, a number of recent research advances can be used to fully address this issue, which can be classified into two categories. The first approach to achieve environment-independent wireless sensing is to extract features that are shared by the behavior data collected under different environments [42], [43]. The second approach is based on transfer learning, which generates synthetic behavior samples or features of the target environment using existing data samples of source environments [44], [45]. Developing a more robust and environment-independent system will be one of our future works.

VII. CONCLUSION

In this paper, we described Gesture-Radar, a novel system to enable natural human-machine interactions. Gesture-Radar uses multiple Doppler radar sensors to capture and recognize arm movements, including gesture types as well as detailed information of the way gestures are performed. In particular, to improve the recognition accuracy, we developed a two-stage classification model and refined the signal proximity matching method. To reduce the environment interference, we proposed to filter the uncorrelated information observed by the dual radars. In addition, we developed a spectrogram analysis method to achieve quantitative gesture detail profiling. Based

on extensive experiments, we demonstrated that Gesture-Radar is able to distinguish various types of gestures with a high accuracy. The study also achieved a high precision for the profiling of fine-grained gesture details, which has significantly advanced the research towards real-world usage.

ACKNOWLEDGMENT

This work is partially supported by the National Major Program for Technological Innovation 2030 - New Generation Artificial Intelligence (No. 2018AAA0100500), the National Natural Science Foundation of China (No. 61960206008, 62072375), and the Fundamental Research Funds for the Central Universities (No. 3102019AX10).

REFERENCES

- [1] L. Ni, Y. Liu, Y. Lau, and A. Patil, "Landmarc: Indoor location sensing using active rfid," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.
- [2] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proc. of MobiCom'14*, 2014, pp. 237–248.
- [3] H. Ketabdar, P. Moghadam, B. Naderi, and M. Roshandel, "Magnetic signatures in air for mobile devices," in *Proc. of MobileHCI'12*. New York, NY, USA: ACM, 2012, pp. 185–188.
- [4] B. Fang, N. Lane, M. Zhang, A. Boran, and F. Kawsar, "Bodyscan: Enabling radio-based sensing on wearable devices for contactless activity and vital sign monitoring," in *Proc. of MobiSys'16*, pp. 97–110.
- [5] D. Balzarotti, M. Cova, and G. Vigna, "Clearshot: Eavesdropping on keyboard input from video," in *Proc. of IEEE S&P'08*, pp. 170–183.
- [6] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE TPAMI*, vol. 19, no. 7, pp. 677–695, 1997.
- [7] M. Stager, P. Lukowicz, and G. Troster, "Implementation and evaluation of a low-power sound-based user activity recognition system," in *Proc. of ACM ISWC'04*, 2004, pp. 138–141.
- [8] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: Using the doppler effect to sense gestures," in *Proc. of CHI'12*, pp. 1911–1914.
- [9] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-fi csi-based behavior recognition: From signals and actions to activities," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 109–115, 2018.
- [10] D. Zhang, H. Wang, and D. Wu, "Toward centimeter-scale human activity sensing with wi-fi signals," *IEEE Computer*, vol. 50, no. 1, pp. 48–57, 2017.
- [11] J. Lien, N. Gillian, M. E. Karagozler, P. Amihoud, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.
- [12] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf, "Leveraging rf-channel fluctuation for activity recognition: Active and passive systems, continuous and rssi-based signal features," in *Proc. of MoMM'13*. New York, NY, USA: ACM, 2013, pp. 43–52.
- [13] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE TMC*, vol. 13, no. 4, pp. 907–920, 2014.
- [14] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Proc. of IEEE INFOCOM'15*, 2015, pp. 1472–1480.
- [15] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" in *Proc. of MobiCom'14*, 2014, pp. 593–604.
- [16] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proc. of MobiCom'15*, 2015, pp. 90–102.
- [17] W. He, K. Wu, Y. Zou, and Z. Ming, "Wig: Wifi-based gesture recognition system," in *Proc. of ICCCN'15*, 2015, pp. 1–7.
- [18] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger: Talk to your smart devices with finger-grained gesture," in *Proc. of ACM UbiComp'16*. New York, NY, USA: ACM, 2016, pp. 250–261.
- [19] M. A. A. Al-qaness and F. Li, "Wiger: Wifi-based gesture recognition system," *ISPRS International Journal of Geo-Information*, vol. 5, no. 6, p. Article No. 92, 2016.
- [20] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures," in *Proc. of MobiCom'14*. ACM, 2014, pp. 617–628.

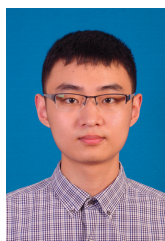
- [21] X. Zheng, J. Wang, L. Shangguan, Z. Zhou, and Y. Liu, "Smokey: Ubiquitous smoking detection with commercial wifi infrastructures," in *Proc. of IEEE INFOCOM'16*, 2016, pp. 1–9.
- [22] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Proc. of CHI'15*. New York, NY, USA: ACM, 2015, pp. 837–846.
- [23] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, 2013.
- [24] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "Widir: Walking direction estimation using wireless signals," in *Proc. of ACM UbiComp'16*. ACM, 2016, pp. 351–362.
- [25] C. Han, K. Wu, Y. Wang, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," in *Proc. of INFOCOM'14*, 2014, pp. 271–279.
- [26] T. Xin, B. Guo, Z. Wang, P. Wang, J. Lam, V. Li, and Z. Yu, "Freesense: A robust approach for indoor human detection using wi-fi signals," *Proc. ACM IMWUT*, vol. 2, no. 3, p. Article 143, 2018.
- [27] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. of MobiCom'13*. New York, NY, USA: ACM, 2013, pp. 27–38.
- [28] Z. Wang, X. Lou, Z. Yu, B. Guo, and X. Zhou, "Enabling non-invasive and real-time human-machine interactions based on wireless sensing and fog computing," *Personal and Ubiquitous Computing*, vol. 23, no. 1, pp. 29–41, 2019.
- [29] T. Ralston, G. Charvat, and J. Peabody, "Real-time through-wall imaging using an ultrawideband multiple-input multiple-output (mimo) phased array radar system," in *Proc. of IEEE ARRAY'10*, 2010, pp. 551–558.
- [30] D.-B. Kim and S.-M. Hong, "Multiple-target tracking and track management for an fmcw radar network," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, p. 159, 2013.
- [31] M. Vossiek, R. Roskosch, and P. Heide, "Precise 3-d object position tracking using fmcw radar," in *Proc. of the 29th European Microwave Conference*, 1999, pp. 234–237.
- [32] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. of FG'2015*, 2015, pp. 1–8.
- [33] Z. Peng, C. Li, J. Muoz-Ferreras, and R. Gomez-Garcia, "An fmcw radar sensor for human gesture recognition in the presence of multiple targets," in *Proc. of IMBIOC'17*, 2017, pp. 1–3.
- [34] K. Chetty, Q. Chen, M. Ritchie, and K. Woodbridge, "A low-cost through-the-wall fmcw radar for stand-off operation and activity detection," *Proc. SPIE*, vol. 10188, p. 1018808, 2017.
- [35] M. A. A. H. Khan, R. Kukkapalli, P. Waradpande, S. Kulandaivel, N. Banerjee, N. Roy, and R. Robucci, "Ram: Radar-based activity monitor," in *Proc. of INFOCOM'16*, 2016, pp. 1–9.
- [36] Z. Li, R. Robucci, N. Banerjee, and C. Patel, "Tongue-n-cheek: Non-contact tongue gesture recognition," in *Proc. of IPSN'15*. New York, NY, USA: ACM, 2015, pp. 95–105.
- [37] S. Dura-Bernal, G. Garreau, C. Andreou, A. Andreou, J. Georgiou, T. Wennekers, and S. Denham, "Human action categorization using ultrasound micro-doppler signatures," in *Human Behavior Understanding*, A. A. Salah and B. Lepri, Eds. Springer, 2011, pp. 18–28.
- [38] F. Wang, M. Tang, Y. Chiu, and T. Horng, "Gesture sensing using retransmitted wireless communication signals based on doppler radar technology," *IEEE TMTT*, vol. 63, no. 12, pp. 4592–4602, 2015.
- [39] X. Lou, Z. Yu, Z. Wang, K. Zhang, and B. Guo, "Gesture-radar: Enabling natural human-computer interactions with radar-based adaptive and robust arm gesture recognition," in *Proc. of IEEE SMC'18*, 2018, pp. 4291–4297.
- [40] V. C. Chen, F. Li, S. Ho, and H. Wechsler, "Micro-doppler effect in radar: phenomenon, model, and simulation study," *IEEE TAES*, vol. 42, no. 1, pp. 2–21, 2006.
- [41] G. Plouffe and A. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE TIM*, vol. 65, no. 2, pp. 305–316, 2016.
- [42] Y. Zheng, Y. Zhang, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Zero-effort cross-domain gesture recognition with wi-fi," in *Proc. of MobiSys'19*, 2019, pp. 313–325.
- [43] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proc. of MobiCom'18*, 2018, pp. 289–304.
- [44] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "Crosssense: Towards cross-site and large-scale wifi sensing," in *Proc. of MobiCom'16*, 2018, pp. 305–320.
- [45] A. Virmani and M. Shahzad, "Position and orientation agnostic gesture recognition using wifi," in *Proc. of MobiSys'17*, 2017, pp. 252–264.



Zhu Wang received the PhD degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China. He is currently an associate professor with Northwestern Polytechnical University. He has worked as a visiting student in the Institut TELECOM SudParis, France, from November 2010 to April 2012. His research interests include pervasive computing, wireless sensing and human-computer interaction. He is a member of the IEEE.



Zhiwen Yu (SM'13) received the PhD degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China. He is currently a professor with Northwestern Polytechnical University. He has worked as an Alexander Von Humboldt Fellow with Mannheim University, Germany, from November 2009 to October 2010. His research interests include pervasive computing and human-computer interaction. He is a senior member of the IEEE.



Xinye Lou received the bachelor's degree from Northwestern Polytechnical University, Xi'an, China. He is a postgraduate student in the School of Computer Science, Northwestern Polytechnical University. His research interests include ubiquitous computing and wireless sensing.



Bin Guo (SM'14) received the PhD degree in computer science from Keio University, Minato, Japan, in 2009. He is currently a professor with Northwestern Polytechnical University, Xi'an, China. He was a postdoctoral researcher in the Institut TELECOM SudParis, France. His research interests include ubiquitous computing, mobile crowd sensing, and human-computer interaction. He is a senior member of the IEEE.



Liming Chen is a professor in the School of Computer Science and Informatics at University of Ulster, Newtownabbey, United Kingdom. He received his B.Eng and M.Eng from Beijing Institute of Technology (BIT), Beijing, China, and his Ph.D in Artificial Intelligence from De Montfort University, UK. His research interests include data analysis, ubiquitous computing, and human-computer interaction. He is a senior member of the IEEE.